ICE-TheOREM - End to end semantically aware eResearch infrastructure for theses

Peter Sefton
University of Southern Queensland
sefton@usq.edu.au

Jim Downing
University of Cambridge
ojd20@cam.ac.uk

2009-05-29

Abstract:

ICE-TheOREM was a project which made several important contributions to the repository domain, promoting deposit by integrating the repository with authoring workflows and enhancing open access, by adding new infrastructure to allow fine-grained embargo management within an institution without impacting on existing open access repository infrastructure.

In the area of scholarly communications workflows, the project produced a complete end-to-end demonstration of eScholarship for word processor users, with tools for authoring, managing and disseminating semantically-rich thesis documents fully integrated with supporting data. This work is focused on theses, as it is well understood that early career researchers are the most likely to lead the charge in new innovations in scholarly publishing and dissemination models.

The authoring tools are built on the ICE content management system, which allows authors to work within a word processing system (as most authors do) with easy-to-use toolbars to structure and format their documents. The ICE system manages both small data files and links to larger data sets. The result is research publication which are available not just as paper-ready PDF files but as fully interactive semantically aware web documents which can be disseminated via repository software such as ePrints, DSpace and Fedora as complete supported web-native **and** PDF publications.

One the technological side, ICE-TheOREM implemented the Object Reuse and Exchange (ORE) protocol to integrate between a content management system, a thesis management system and multiple repository software packages and looked at ways to describe aggregate objects which include both data and documents, which can be generalized to domains other than chemistry. ICE-TheOREM has demonstrated how focusing on the use of the web architecture (including ORE) enables repository functions to be distributed between systems for complex, data-rich compound objects.

Introduction

In this presentation we will demonstrate several contributions that the project has made. We begin by looking at the project in terms of workflows, and the impact on users, and then go on to look at the technical architecture that supports it.

Innovations in workflow

The TheOREM project aimed to exercise the OAI-ORE protocol (IONSREPORT 2008) in the context of chemical theses – with content . The ICE extension to that project showed how chemical theses could be authored in a word processing environment. We have been able to demonstrate theses that are both Supported in Neylon's terms (Neylon 2008) and datuments (Murray-Rust & Rzepa 2004) that is they are hypertext aggregations of document and data, which are both human and machine-readable.

Other innovations in the workflow/lifecyle of a thesis include:

- Effective capture of metadata (technical and descriptive) as part of the authoring process rather as part of deposit process. In fact, the deposit process has been replaced altogether.
- Showing how repository ingest can be made a by-product of an existing workflow, with data moving between systems based on the functional requirements of the stakeholders rather than a mandate to deposit data and papers. We contend that this direction whereby services are driven by the immediate motivations of the participants will be easier and quicker to bootstrap to a sustainable long-term business model than those driven by edict.

Technical Architecture

- Motivating capture of data as part of writing process by automating some of the currently mundane but time-consuming tasks involved in preparing data for presentation in publication, without destroying the data in the process.
- Ready-to-use, semantically rich authoring tools including Microsoft Word and OpenOffice.org templates with an emphasis at this stage on chemistry, but with potential for use in other disciplines.
- Working implementations of ORE – including code to both push content using SWORD and harvest it using the ATOM archive format which may be reused in other projects. This is achieved using metadata construction 'invisible' to the author, who is guided into creating good metadata and data through intuitive extensions to a familiar interface.
- A proof-of-concept repository architecture for start-to-finish thesis management from authoring to dissemination, with an innovative approach to embargo management based on OpenID. This includes a nascent thesis repository built on Fedora-commons and The Fascinator (a Fedora front-end).

Further work required

The work reported here is a proof of principle for the ORE technology and a first step towards larger scale trials of repository-integrated thesis authoring workflows. A PhD thesis takes years to complete, so a true test of this infrastructure will involve a long term commitment. This commitment is being made at the Australian Digital Futures Institute – beginning in early 2009 all the theses begin completed by institute staff and affiliates will be housed in a system derived from the TheOREM

work (Observatory PASCAL http://www.obs-pascal.com/).

Acknowledgements

ICE-Theorem was a joint project between the University of Cambridge (UC) and the University of Southern Queensland (USQ) funded by the JISC (Jacobs 2008)

References

IONSREPORT, S., 2008. OAI-ORE specifications. *Scholarly Communications Report*, 12(1), 5-5.

Jacobs, N., 2008. Small-Scale OAI Object Re-Use and Exchange Experiments. Available at: http://www.jisc.ac.uk/whatwedo/programmes/digitalrepositories2007/oaioredemonstrators.aspx [Accessed February 5, 2009].

Murray-Rust, P. & Rzepa, H.S., 2004. The Next Big Thing: From Hypermedia to Datuments. *Journal of Digital Information*, 5(1), 248.

Neylon, C., 2008. Science in the open » A personal view of Open Science - Part IV - Policies and standards. Available at: http://blog.openwetware.org/scienceintheopen/2008/10/26/a-personal-view-of-open-science-part-iv-policies-and-standards/ [Accessed February 5, 2009].